

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-268663

(43)Date of publication of application : 20.09.2002

(51)Int.Cl. G10L 13/08
 A63H 11/00
 G10L 13/00
 G10L 13/06
 // B25J 13/00

(21)Application number : 2001-065072

(71)Applicant : SONY CORP

(22)Date of filing : 08.03.2001

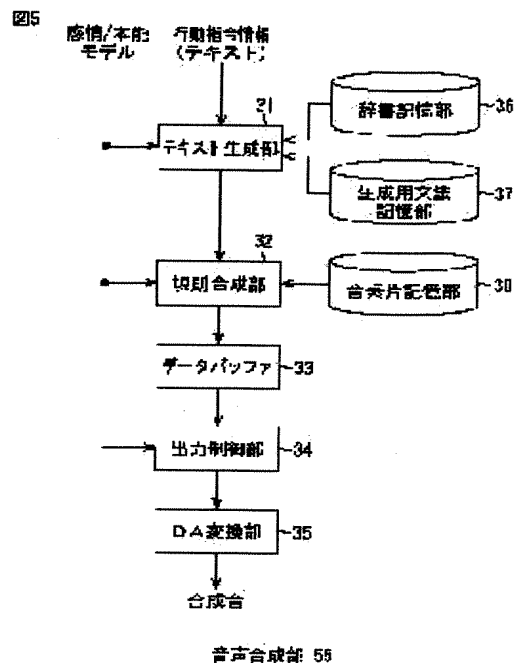
(72)Inventor : ASANO KOJI
 KOBAYASHI KENICHIRO
 YAMAZAKI NOBUHIDE
 KARIYA SHINICHI
 FUJITA YAEKO

(54) VOICE SYNTHESIZER, VOICE SYNTHESIS METHOD, PROGRAM AND RECORDING MEDIUM

(57)Abstract:

PROBLEM TO BE SOLVED: To actualize a pet robot, etc., having high interactivity.

SOLUTION: A text generation part 31 and a rule synthesis part 32 generate synthesized voice data corresponding to a text included in action command information according to the action command information and the data are stored in a data buffer 33. An output control part 34, on the other hand, controls the output of the synthesized voice data stored in the data buffer 33 according to the state of the feeling of the pet robot.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision
of rejection]

[Date of requesting appeal against examiner's
decision of rejection]

[Date of extinction of right]

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号
特開2002-268663
(P2002-268663A)

(43) 公開日 平成14年9月20日 (2002.9.20)

(51) Int.Cl. ⁷	識別記号	F I	テーマコード [*] (参考)
G 1 0 L 13/08		A 6 3 H 11/00	Z 2 C 1 5 0
A 6 3 H 11/00		B 2 5 J 13/00	Z 3 C 0 0 7
G 1 0 L 13/00		C 1 0 L 3/00	H 5 D 0 4 5
13/06			Q
// B 2 5 J 13/00		5/04	F

審査請求 未請求 請求項の数 8 O L (全 13 頁)

(21) 出願番号 特願2001-65072(P2001-65072)

(22) 出願日 平成13年3月8日 (2001.3.8)

(71) 出願人 000002185

ソニー株式会社

東京都品川区北品川 6 丁目 7 番35号

(72) 発明者 浅野 康治

東京都品川区北品川 6 丁目 7 番35号 ソニー株式会社内

(72) 発明者 小林 賢一郎

東京都品川区北品川 6 丁目 7 番35号 ソニー株式会社内

(74) 代理人 100082131

弁理士 稲本 義雄

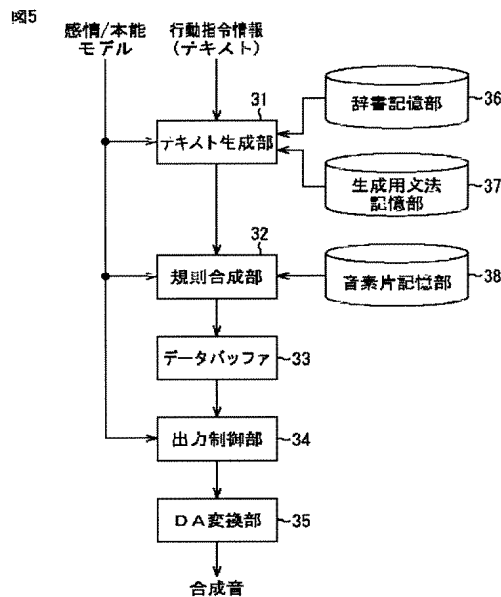
最終頁に続く

(54) 【発明の名称】 音声合成装置および音声合成方法、並びにプログラムおよび記録媒体

(57) 【要約】

【課題】 インタラクティブ性の高いペットロボット等を実現する。

【解決手段】 テキスト生成部 31 および規則合成部 32 では、行動指令情報にしたがい、その行動指令情報に含まれるテキストに対応する合成音データが生成され、データバッファ 33 に記憶される。一方、出力制御部 34 は、ペットロボットの感情の状態に基づき、データバッファ 33 に記憶された合成音データの出力を制御する。



音声合成部 55

【特許請求の範囲】

【請求項1】 情報処理装置の制御にしたがって、合成音を生成する音声合成装置であって、

前記情報処理装置の制御にしたがって、合成音を生成する音声合成手段と、

前記情報処理装置の内部状態に基づいて、前記合成音の出力を制御する出力制御手段とを備えることを特徴とする音声合成装置。

【請求項2】 前記出力制御手段は、前記情報処理装置の内部状態に基づいて、前記合成音の出力を停止することを特徴とする音声合成装置。

【請求項3】 前記音声合成手段は、前記出力制御手段が前記合成音の出力を停止した後、前記情報処理装置の内部状態に基づいて、前記合成音を再生成し、

前記出力制御手段は、前記音声合成手段において再生成された前記合成音を出力することを特徴とする請求項2に記載の音声合成装置。

【請求項4】 前記情報処理装置は、実際の、または仮想的なロボットであることを特徴とする請求項1に記載の音声合成装置。

【請求項5】 前記出力制御手段は、前記ロボットの感情または本能の状態に基づいて、前記合成音の出力を制御することを特徴とする請求項4に記載の音声合成装置。

【請求項6】 情報処理装置の制御にしたがって、合成音を生成する音声合成方法であって、
前記情報処理装置の制御にしたがって、合成音を生成する音声合成ステップと、
前記情報処理装置の内部状態に基づいて、前記合成音の出力を制御する出力制御ステップとを備えることを特徴とする音声合成方法。

【請求項7】 情報処理装置の制御にしたがって、合成音を生成する音声合成処理を、コンピュータに行わせる

{ (' /) P3 (' /) . . . }

【0005】(1)においては、発音記号をカタカナで表している。また、タグ'は、アクセントを表し、タグ{}および()は、フレーズの区切りを表す。さらに、タグP3のうちのPは、ポーズを表し、続く数字の3は、ポーズの長さを表す。また、タグ/は、アクセント句の区切りを表す。

【0006】(1)の発音記号によれば、音声合成装置では、音韻が「これまで進められた研究は、大きな成果を・・・」という合成音が生成される。

【0007】

【発明が解決しようとする課題】ところで、最近、例えば、ペット型のペットロボット等として、音声合成装置を搭載し、ユーザに話しかけたり、ユーザと会話(対話)を行うものが提案されている。

【0008】さらに、ペットロボットとしては、感情の

プログラムであって、

前記情報処理装置の制御にしたがって、合成音を生成する音声合成ステップと、

前記情報処理装置の内部状態に基づいて、前記合成音の出力を制御する出力制御ステップとを備えることを特徴とするプログラム。

【請求項8】 情報処理装置の制御にしたがって、合成音を生成する音声合成処理を、コンピュータに行わせるプログラムが記録されている記録媒体であって、

前記情報処理装置の制御にしたがって、合成音を生成する音声合成ステップと、

前記情報処理装置の内部状態に基づいて、前記合成音の出力を制御する出力制御ステップとを備えるプログラムが記録されていることを特徴とする記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、音声合成装置および音声合成方法、並びにプログラムおよび記録媒体に関し、特に、例えば、エンタテインメント用のロボットの感情等の状態に基づいて、合成音出力を制御するようにすることで、インタラクティブ性の高いロボットを実現すること等ができるようにする音声合成装置および音声合成方法、並びにプログラムおよび記録媒体に関する。

【0002】

【従来の技術】従来の音声合成装置においては、テキスト、またはそのテキストを解析して得られる発音記号に基づいて、合成音が生成される。また、音声合成装置では、例えば、合成音の発話速度や、高さ、強さ、ポーズの長さ等が、音声合成装置に入力されるテキストや発音記号に挿入されたタグに基づいて制御される。

【0003】ここで、タグが挿入された発音記号としては、例えば、次のようなものがある。

【0004】

・・・(1)

状態を表す感情モデルを取り入れ、その感情モデルが表す感情の状態に応じて、合成音の発話速度や、高さ、強さ、ポーズの長さ等を制御し、感情を表現した合成音を出力するものも提案されている。

【0009】なお、音声に含まれる発話意図や感情と、発話速度や基本周波数等との関係については、例えば、前川、「音声によるパラ言語情報の伝達：言語学の立場から」、日本音響学会平成9年度秋季研究発表会講演論文集、pp.381-384(以下、文献1という)等に記載されている。

【0010】ところで、例えば、上述のような音声合成装置を搭載したペットロボットにおいて、ユーザとのインタラクティブ性を向上させるためには、合成音を出力している最中に、感情の状態の変化に応じて、合成音の出力を停止したり、合成音の発話速度や高さ等を変更す

ることができるようにするのが望ましい。

【0011】しかしながら、従来の音声合成装置では、上述したようなタグが挿入された発音記号列にしたがって、合成音が生成されるため、合成音が出力されている最中に、感情の状態が変化した場合に、リアルタイムで、合成音の出力を停止したりすること等が困難であった。

【0012】本発明は、このような状況に鑑みてなされたものであり、ロボットの感情等の状態に応じて、合成音の出力を制御することにより、インタラクティブ性の高いロボット等を実現することができるようにするものである。

【0013】

【課題を解決するための手段】本発明の音声合成装置は、情報処理装置の制御にしたがって、合成音を生成する音声合成手段と、情報処理装置の内部状態に基づいて、合成音の出力を制御する出力制御手段とを備えることを特徴とする。

【0014】本発明の音声合成方法は、情報処理装置の制御にしたがって、合成音を生成する音声合成ステップと、情報処理装置の内部状態に基づいて、合成音の出力を制御する出力制御ステップとを備えることを特徴とする。

【0015】本発明のプログラムは、情報処理装置の制御にしたがって、合成音を生成する音声合成ステップと、情報処理装置の内部状態に基づいて、合成音の出力を制御する出力制御ステップとを備えることを特徴とする。

【0016】本発明の記録媒体は、情報処理装置の制御にしたがって、合成音を生成する音声合成ステップと、情報処理装置の内部状態に基づいて、合成音の出力を制御する出力制御ステップとを備えるプログラムが記録されていることを特徴とする。

【0017】本発明の音声合成装置および音声合成方法、並びにプログラムにおいては、情報処理装置の制御にしたがって、合成音が生成される一方、情報処理装置の内部状態に基づいて、合成音の出力が制御される。

【0018】

【発明の実施の形態】図1は、本発明を適用したロボットの一実施の形態の外観構成例を示しており、図2は、その電氣的構成例を示している。

【0019】本実施の形態では、ロボットは、例えば、犬等の四つ足の動物の形状のものとなっており、胴体部ユニット2の前後左右に、それぞれ脚部ユニット3A、3B、3C、3Dが連結されるとともに、胴体部ユニット2の前端部と後端部に、それぞれ頭部ユニット4と尻尾部ユニット5が連結されることにより構成されている。

【0020】尻尾部ユニット5は、胴体部ユニット2の上面に設けられたベース部5Bから、2自由度をもって

湾曲または揺動自在に引き出されている。

【0021】胴体部ユニット2には、ロボット全体の制御を行うコントローラ10、ロボットの動力源となるバッテリー11、並びにバッテリーセンサ12および熱センサ13からなる内部センサ部14などが収納されている。

【0022】頭部ユニット4には、「耳」に相当するマイク（マイクロフォン）15、「目」に相当するCCD（Charge Coupled Device）カメラ16、触覚に相当するタッチセンサ17、「口」に相当するスピーカ18などが、それぞれ所定位置に配設されている。また、頭部ユニット4には、口の下顎に相当する下顎部4Aが1自由度をもって可動に取り付けられており、この下顎部4Aが動くことにより、ロボットの口の開閉動作が実現されるようになっている。

【0023】脚部ユニット3A乃至3Dそれぞれの関節部分や、脚部ユニット3A乃至3Dそれぞれと胴体部ユニット2の連結部分、頭部ユニット4と胴体部ユニット2の連結部分、頭部ユニット4と下顎部4Aの連結部分、並びに尻尾部ユニット5と胴体部ユニット2の連結部分などには、図2に示すように、それぞれアクチュエータ3AA₁乃至3AA_K、3BA₁乃至3BA_K、3CA₁乃至3CA_K、3DA₁乃至3DA_K、4A₁乃至4A_L、5A₁および5A₂が配設されている。

【0024】頭部ユニット4におけるマイク15は、ユーザからの発話を含む周囲の音声（音）を集音し、得られた音声信号を、コントローラ10に送出する。CCDカメラ16は、周囲の状況を撮像し、得られた画像信号を、コントローラ10に送出する。

【0025】タッチセンサ17は、例えば、頭部ユニット4の上部に設けられており、ユーザからの「なでる」や「たたく」といった物理的な働きかけにより受けた圧力を検出し、その検出結果を圧力検出信号としてコントローラ10に送出する。

【0026】胴体部ユニット2におけるバッテリーセンサ12は、バッテリー11の残量を検出し、その検出結果を、バッテリー残量検出信号としてコントローラ10に送出する。熱センサ13は、ロボット内部の熱を検出し、その検出結果を、熱検出信号としてコントローラ10に送出する。

【0027】コントローラ10は、CPU（Central Processing Unit）10Aやメモリ10B等を内蔵しており、CPU10Aにおいて、メモリ10Bに記憶された制御プログラムが実行されることにより、各種の処理を行う。

【0028】即ち、コントローラ10は、マイク15や、CCDカメラ16、タッチセンサ17、バッテリーセンサ12、熱センサ13から与えられる音声信号、画像信号、圧力検出信号、バッテリー残量検出信号、熱検出信号に基づいて、周囲の状況や、ユーザからの指令、ユーザからの働きかけなどの有無を判断する。

【0029】さらに、コントローラ10は、この判断結果等に基づいて、続く行動を決定し、その決定結果に基づいて、アクチュエータ3A_{A1}乃至3A_{A_K}、3B_{A1}乃至3B_{A_K}、3C_{A1}乃至3C_{A_K}、3D_{A1}乃至3D_{A_K}、4A₁乃至4A_L、5A₁、5A₂のうちの必要なものを駆動させる。これにより、頭部ユニット4を上下左右に振らせたり、下顎部4Aを開閉させる。さらには、尻尾部ユニット5を動かしたり、各脚部ユニット3A乃至3Dを駆動して、ロボットを歩行させるなどの行動を行わせる。

【0030】また、コントローラ10は、必要に応じて、合成音を生成し、スピーカ18に供給して出力させたり、ロボットの「目」の位置に設けられた図示しないLED (Light Emitting Diode) を点灯、消灯または点滅させる。

【0031】以上のようにして、ロボットは、周囲の状況等に基づいて自律的に行動をとるようになっている。

【0032】次に、図3は、図2のコントローラ10の機能的構成例を示している。なお、図3に示す機能的構成は、CPU10Aが、メモリ10Bに記憶された制御プログラムを実行することで実現されるようになっている。

【0033】コントローラ10は、特定の外部状態を認識するセンサ入力処理部50、センサ入力処理部50の認識結果を累積して、感情や、本能、成長の状態を表現するモデル記憶部51、センサ入力処理部50の認識結果等に基づいて、続く行動を決定する行動決定機構部52、行動決定機構部52の決定結果に基づいて、実際にロボットに行動を起こさせる姿勢遷移機構部53、各アクチュエータ3A_{A1}乃至5A₁および5A₂を駆動制御する制御機構部54、並びに合成音を生成する音声合成部55から構成されている。

【0034】センサ入力処理部50は、マイク15や、CCDカメラ16、タッチセンサ17等から与えられる音声信号、画像信号、圧力検出信号等に基づいて、特定の外部状態や、ユーザからの特定の働きかけ、ユーザからの指示等を認識し、その認識結果を表す状態認識情報を、モデル記憶部51および行動決定機構部52に通知する。

【0035】即ち、センサ入力処理部50は、音声認識部50Aを有しており、音声認識部50Aは、マイク15から与えられる音声信号について音声認識を行う。そして、音声認識部50Aは、その音声認識結果としての、例えば、「歩け」、「伏せ」、「ボールを追いかける」等の指令その他を、状態認識情報として、モデル記憶部51および行動決定機構部52に通知する。

【0036】また、センサ入力処理部50は、画像認識部50Bを有しており、画像認識部50Bは、CCDカメラ16から与えられる画像信号を用いて、画像認識処理を行う。そして、画像認識部50Bは、その処理の結

果、例えば、「赤い丸いもの」や、「地面に対して垂直なかつ所定高さ以上の平面」等を検出したときには、「ボールがある」や、「壁がある」等の画像認識結果を、状態認識情報として、モデル記憶部51および行動決定機構部52に通知する。

【0037】さらに、センサ入力処理部50は、圧力処理部50Cを有しており、圧力処理部50Cは、タッチセンサ17から与えられる圧力検出信号を処理する。そして、圧力処理部50Cは、その処理の結果、所定の閾値以上で、かつ短時間の圧力を検出したときには、「たたかれた(しかられた)」と認識し、所定の閾値未満で、かつ長時間の圧力を検出したときには、「なでられた(ほめられた)」と認識して、その認識結果を、状態認識情報として、モデル記憶部51および行動決定機構部52に通知する。

【0038】モデル記憶部51は、ロボットの感情、本能、成長の状態を表現する感情モデル、本能モデル、成長モデルをそれぞれ記憶、管理している。

【0039】ここで、感情モデルは、例えば、「うれしさ」、「悲しさ」、「怒り」、「楽しさ」等の感情の状態(度合い)を、所定の範囲(例えば、-1.0乃至1.0等)の値によってそれぞれ表し、センサ入力処理部50からの状態認識情報や時間経過等に基づいて、その値を変化させる。本能モデルは、例えば、「食欲」、「睡眠欲」、「運動欲」等の本能による欲求の状態(度合い)を、所定の範囲の値によってそれぞれ表し、センサ入力処理部50からの状態認識情報や時間経過等に基づいて、その値を変化させる。成長モデルは、例えば、「幼年期」、「青年期」、「熟年期」、「老年期」等の成長の状態(度合い)を、所定の範囲の値によってそれぞれ表し、センサ入力処理部50からの状態認識情報や時間経過等に基づいて、その値を変化させる。

【0040】モデル記憶部51は、上述のようにして感情モデル、本能モデル、成長モデルの値で表される感情、本能、成長の状態を、状態情報として、行動決定機構部52に送出する。

【0041】なお、モデル記憶部51には、センサ入力処理部50から状態認識情報が供給される他、行動決定機構部52から、ロボットの現在または過去の行動、具体的には、例えば、「長時間歩いた」などの行動の内容を示す行動情報が供給されるようになっており、モデル記憶部51は、同一の状態認識情報が与えられても、行動情報が示すロボットの行動に応じて、異なる状態情報を生成するようになっている。

【0042】即ち、例えば、ロボットが、ユーザに挨拶をし、ユーザに頭を撫でられた場合には、ユーザに挨拶をしたという行動情報と、頭を撫でられたという状態認識情報とが、モデル記憶部51に与えられ、この場合、モデル記憶部51では、「うれしさ」を表す感情モデルの値が増加される。

【0043】一方、ロボットが、何らかの仕事を実行中に頭を撫でられた場合には、仕事を実行中であるという行動情報と、頭を撫でられたという状態認識情報とが、モデル記憶部51に与えられ、この場合、モデル記憶部51では、「うれしさ」を表す感情モデルの値は変化されない。

【0044】このように、モデル記憶部51は、状態認識情報だけでなく、現在または過去のロボットの行動を示す行動情報も参照しながら、感情モデルの値を設定する。これにより、例えば、何らかのタスクを実行中に、ユーザが、いたずらするつもりで頭を撫でたときに、「うれしさ」を表す感情モデルの値を増加させるような、不自然な感情の変化が生じることを回避することができる。

【0045】なお、モデル記憶部51は、本能モデルおよび成長モデルについても、感情モデルにおける場合と同様に、状態認識情報および行動情報の両方に基づいて、その値を増減させるようになっている。また、モデル記憶部51は、感情モデル、本能モデル、成長モデルそれぞれの値を、他のモデルの値にも基づいて増減させるようになっている。

【0046】行動決定機構部52は、センサ入力処理部50からの状態認識情報や、モデル記憶部51からの状態情報、時間経過等に基づいて、次の行動を決定し、決定された行動の内容を、行動指令情報として、姿勢遷移機構部53に送出する。

【0047】即ち、行動決定機構部52は、ロボットがとり得る行動をステート（状態）(state)に対応させた有限オートマンを、ロボットの行動を規定する行動モデルとして管理しており、この行動モデルとしての有限オートマンにおけるステートを、センサ入力処理部50からの状態認識情報や、モデル記憶部51における感情モデル、本能モデル、または成長モデルの値、時間経過等に基づいて遷移させ、遷移後のステートに対応する行動を、次にとるべき行動として決定する。

【0048】ここで、行動決定機構部52は、所定の取りが(trigger)があったことを検出すると、ステートを遷移させる。即ち、行動決定機構部52は、例えば、現在のステートに対応する行動を実行している時間が所定時間に達したときや、特定の状態認識情報を受信したとき、モデル記憶部51から供給される状態情報が示す感情や、本能、成長の状態の値が所定の閾値以下または以上になったとき等に、ステートを遷移させる。

【0049】なお、行動決定機構部52は、上述したように、センサ入力処理部50からの状態認識情報だけでなく、モデル記憶部51における感情モデルや、本能モデル、成長モデルの値等に基づいて、行動モデルにおけるステートを遷移させることから、同一の状態認識情報が入力されても、感情モデルや、本能モデル、成長モデルの値（状態情報）によっては、ステートの遷移先は異

なるものとなる。

【0050】その結果、行動決定機構部52は、例えば、状態情報が、「怒っていない」こと、および「お腹がすいていない」ことを表している場合において、状態認識情報が、「目の前に手のひらが差し出された」ことを表しているときには、目の前に手のひらが差し出されたことに応じて、「お手」という行動をとらせる行動指令情報を生成し、これを、姿勢遷移機構部53に送出する。

【0051】また、行動決定機構部52は、例えば、状態情報が、「怒っていない」こと、および「お腹がすいている」ことを表している場合において、状態認識情報が、「目の前に手のひらが差し出された」ことを表しているときには、目の前に手のひらが差し出されたことに応じて、「手のひらをべろべろなめる」ような行動を行わせるための行動指令情報を生成し、これを、姿勢遷移機構部53に送出する。

【0052】また、行動決定機構部52は、例えば、状態情報が、「怒っている」ことを表している場合において、状態認識情報が、「目の前に手のひらが差し出された」ことを表しているときには、状態情報が、「お腹がすいている」ことを表していても、また、「お腹がすいていない」ことを表していても、「おいと横を向く」ような行動を行わせるための行動指令情報を生成し、これを、姿勢遷移機構部53に送出する。

【0053】なお、行動決定機構部52には、モデル記憶部51から供給される状態情報が示す感情や、本能、成長の状態に基づいて、遷移先のステートに対応する行動のパラメータとしての、例えば、歩行の速度や、手足を動かす際の動きの大きさおよび速度などを決定させることができ、この場合、それらのパラメータを含む行動指令情報が、姿勢遷移機構部53に送出される。

【0054】また、行動決定機構部52では、上述したように、ロボットの頭部や手足等を動作させる行動指令情報の他、ロボットに発話を行わせる行動指令情報も生成される。ロボットに発話を行わせる行動指令情報は、音声合成部55に供給されるようになっており、音声合成部55に供給される行動指令情報には、音声合成部55に生成させる合成音に対応するテキスト等が含まれる。そして、音声合成部55は、行動決定部52から行動指令情報を受信すると、その行動指令情報に含まれるテキストに基づき、合成音を生成し、スピーカ18に供給して出力させる。これにより、スピーカ18からは、例えば、ロボットの鳴き声、さらには、「お腹がすいた」等のユーザへの各種の要求、「何？」等のユーザの呼びかけに対する応答その他の音声出力が行われる。ここで、音声合成部55には、モデル記憶部51から状態情報も供給されるようになっており、音声合成部55は、この状態情報が示す感情の状態に基づいて韻律等を制御した合成音を生成することが可能となっている。な

お、音声合成部55では、感情の他、本能や本能の状態に基づいて韻律等を制御した合成音を生成することも可能である。また、行動決定機構部52は、合成音を出力する場合には、下顎部4Aを開閉させる行動指令情報を、必要に応じて生成し、姿勢遷移機構部53に出力する。この場合、合成音の出力に同期して、下顎部4Aが開閉し、ユーザに、ロボットがしゃべっているかのような印象を与えることができる。

【0055】姿勢遷移機構部53は、行動決定機構部52から供給される行動指令情報に基づいて、ロボットの姿勢を、現在の姿勢から次の姿勢に遷移させるための姿勢遷移情報を生成し、これを制御機構部54に送出する。

【0056】ここで、現在の姿勢から次に遷移可能な姿勢は、例えば、胴体や手や足の形状、重さ、各部の結合状態のようなロボットの物理的形状と、関節が曲がる方向や角度のようなアクチュエータ3AA₁乃至5A₁および5A₂の機構とによって決定される。

【0057】また、次の姿勢としては、現在の姿勢から直接遷移可能な姿勢と、直接には遷移できない姿勢とがある。例えば、4本足のロボットは、手足を大きく投げ出して寝転んでいる状態から、伏せた状態へ直接遷移することはできるが、立った状態へ直接遷移することはできず、一旦、手足を胴体近くに引き寄せて伏せた姿勢になり、それから立ち上がるという2段階の動作が必要である。また、安全に実行できない姿勢も存在する。例えば、4本足のロボットは、その4本足で立っている姿勢から、両前足を挙げてバンザイをしようとすると、簡単に転倒してしまう。

【0058】このため、姿勢遷移機構部53は、直接遷移可能な姿勢をあらかじめ登録しておき、行動決定機構部52から供給される行動指令情報が、直接遷移可能な姿勢を示す場合には、その行動指令情報を、そのまま姿勢遷移情報として、制御機構部54に送出する。一方、行動指令情報が、直接遷移不可能な姿勢を示す場合には、姿勢遷移機構部53は、遷移可能な他の姿勢に一旦遷移した後に、目的の姿勢まで遷移させるような姿勢遷移情報を生成し、制御機構部54に送出する。これによりロボットが、遷移不可能な姿勢を無理に実行しようとする事態や、転倒するような事態を回避することができるようにしている。

【0059】制御機構部54は、姿勢遷移機構部53からの姿勢遷移情報にしたがって、アクチュエータ3AA₁乃至5A₁および5A₂を駆動するための制御信号を生成し、これを、アクチュエータ3AA₁乃至5A₁および5A₂に送出する。これにより、アクチュエータ3AA₁乃至5A₁および5A₂は、制御信号にしたがって駆動し、ロボットは、自律的に行動を起こす。

【0060】次に、図4は、図3の音声認識部50Aの構成例を示している。

【0061】マイク15からの音声信号は、AD(Analog Digital)変換部21に供給される。AD変換部21では、マイク15からのアナログ信号である音声信号がサンプリング、量子化され、ディジタル信号である音声データにA/D変換される。この音声データは、特徴抽出部22および音声区間検出部27に供給される。

【0062】特徴抽出部22は、そこに入力される音声データについて、適当なフレームごとに、例えば、MFCC(Mel Frequency Cepstrum Coefficient)分析を行い、その分析の結果得られるMFCCを、特徴パラメータ(特徴ベクトル)として、マッチング部23に出力する。なお、特徴抽出部22では、その他、例えば、線形予測係数、ケプストラム係数、線スペクトル対、所定の周波数帯域ごとのパワー(フィルタバンクの出力)等を、特徴パラメータとして抽出することが可能である。

【0063】マッチング部23は、特徴抽出部22からの特徴パラメータを用いて、音響モデル記憶部24、辞書記憶部25、および文法記憶部26を必要に応じて参照しながら、マイク15に入力された音声(入力音声)を、例えば、連続分布HMM(Hidden Markov Model)法に基づいて音声認識する。

【0064】即ち、音響モデル記憶部24は、音声認識する音声の言語における個々の音素や音節などの音響的な特徴を表す音響モデルを記憶している。ここでは、連続分布HMM法に基づいて音声認識を行うので、音響モデルとしては、HMM(Hidden Markov Model)が用いられる。辞書記憶部25は、認識対象の各単語について、その発音に関する情報(音韻情報)が記述された単語辞書を記憶している。文法記憶部26は、辞書記憶部25の単語辞書に登録されている各単語が、どのように連鎖する(つながる)かを記述した文法規則を記憶している。ここで、文法規則としては、例えば、文脈自由文法(CFG)や、統計的な単語連鎖確率(N-gram)などに基づく規則を用いることができる。

【0065】マッチング部23は、辞書記憶部25の単語辞書を参照することにより、音響モデル記憶部24に記憶されている音響モデルを接続することで、単語の音響モデル(単語モデル)を構成する。さらに、マッチング部23は、幾つかの単語モデルを、文法記憶部26に記憶された文法規則を参照することにより接続し、そのようにして接続された単語モデルを用いて、特徴パラメータに基づき、連続分布HMM法によって、マイク15に入力された音声を認識する。即ち、マッチング部23は、特徴抽出部22が出力する時系列の特徴パラメータが観測されるスコア(尤度)が最も高い単語モデルの系列を検出し、その単語モデルの系列に対応する単語列の音韻情報(読み)を、音声の認識結果として出力する。

【0066】より具体的には、マッチング部23は、接続された単語モデルに対応する単語列について、各特徴パラメータの出現確率(出力確率)を累積し、その累積

値をスコアとして、そのスコアを最も高くする単語列の音韻情報を、音声認識結果として出力する。

【0067】以上のようにして出力される、マイク15に入力された音声の認識結果は、状態認識情報として、モデル記憶部51および行動決定機構部52に出力される。

【0068】なお、音声区間検出部27は、AD変換部21からの音声データについて、特徴抽出部22がMFCC分析を行うのと同様のフレームごとに、例えば、パワーを算出している。さらに、音声区間検出部27は、各フレームのパワーを、所定の閾値と比較し、その閾値以上のパワーを有するフレームで構成される区間を、ユーザの音声が入力されている音声区間として検出する。そして、音声区間検出部27は、検出した音声区間を、特徴抽出部22とマッチング部23に供給しており、特徴抽出部22とマッチング部23は、音声区間のみを対象に処理を行う。

【0069】次に、図5は、図3の音声合成部55の構成例を示している。

【0070】テキスト生成部31には、行動決定機構部52が出力する、音声合成の対象とするテキストを含む行動指令情報が供給されるようになっており、テキスト生成部31は、辞書記憶部36や生成用文法記憶部37を参照しながら、その行動指令情報に含まれるテキストを解析する。

【0071】即ち、辞書記憶部36には、各単語の品詞情報や、読み、アクセント等の情報が記述された単語辞書が記憶されており、また、生成用文法記憶部37には、辞書記憶部36の単語辞書に記述された単語について、単語連鎖に関する制約等の文法規則が記憶されている。そして、テキスト生成部31は、この単語辞書および文法規則に基づいて、そこに入力されるテキストの形態素解析や構文解析等の解析を行い、後段の規則合成部32で行われる規則音声合成に必要な情報を抽出する。ここで、規則音声合成に必要な情報としては、例えば、ポーズの位置や、アクセントおよびイントネーションを制御するための情報その他の韻律情報や、各単語の発音等の音韻情報などがある。

【0072】テキスト生成部31で得られた情報は、規則合成部32に供給され、規則合成部32は、音素片記憶部38を用いて、テキスト生成部31に入力されたテキストに対応する合成音の音声データ（デジタルデータ）を生成する。

【0073】即ち、音素片記憶部38には、例えば、CV(Consonant, Vowel)や、VCV、CVC等の形で音素片データが記憶されており、規則合成部32は、テキスト生成部31からの情報に基づいて、必要な音素片データを接続し、さらに、ポーズ、アクセント、イントネーション等を適切に付加することで、テキスト生成部31に入力されたテキストに対応する合成音データを生成す

る。

【0074】この音声データは、データバッファ33に供給される。データバッファ33は、規則合成部32から供給される合成音データを記憶する。

【0075】出力制御部34は、定期的または不定期に、モデル記憶部51（図3）に記憶された感情モデル等をチェックし、その感情モデル等に基づいて、データバッファ33に記憶された合成音データの出力を制御する。

【0076】即ち、出力制御部34は、感情モデルの値（感情モデル値）が、ある条件を満たすとき、データバッファ33に記憶された合成音データを読み出し、DA(Digital Analogue)変換部35に供給する。この場合、DA変換部35は、ディジタル信号としての合成音データを、アナログ信号としての音声信号にD/A変換する。この音声信号は、スピーカ18に供給され、これにより、テキスト生成部31に入力されたテキストに対応する合成音出力される。

【0077】また、出力制御部34は、感情モデル値が、他の条件を満たすとき、データバッファ33に記憶された合成音データの読み出しを停止する。この場合、スピーカ18からの合成音の出力は停止する。

【0078】また、出力制御部34は、感情モデル値が、さらに他の条件を満たすとき、データバッファ33に記憶された合成音データの読み出しを停止し、その後、データバッファ33に記憶された合成音データの読み出しを再開する。この場合、スピーカ18からの合成音の出力は、一旦停止され、その後、再開される。

【0079】なお、テキスト生成部31および規則合成部32も、出力制御部34と同様に、モデル記憶部51（図3）に記憶された感情モデルの値（感情モデル値）や本能モデルの値（本能モデル値）をチェックするようになっており、この感情モデル値や本能モデル値を考慮して処理を行うようになっている。

【0080】次に、図6のフローチャートを参照して、図5の音声合成部55による音声合成処理について説明する。

【0081】行動決定機構部52が、音声合成の対象とするテキストを含む行動指令情報を、音声合成部55に出力すると、テキスト生成部31は、ステップS1において、その行動指令情報を受信し、ステップS2に進む。ステップS2では、テキスト生成部31および規則合成部32において、モデル記憶部51を参照することで、感情モデル値や本能モデル値が認識（チェック）され、ステップS3に進む。

【0082】ステップS3では、テキスト生成部31において、行動決定機構部52からの行動指令情報に含まれるテキストから、実際に合成音として出力するテキスト（以下、適宜、発話テキストという）を生成する際に用いる語彙（発話語彙）が、感情モデル値や本能モデル

値に基づいて設定され、ステップS4に進む。ステップS4では、テキスト生成部31において、ステップS3で設定された発話語彙を用いて、行動指令情報に含まれるテキストに対応する発話テキストが生成される。

【0083】即ち、行動決定機構部52からの行動指令情報に含まれるテキストは、例えば、標準的な感情および本能の状態における発話を前提としたものとなっており、ステップS4では、そのテキストが、ロボットの感情や本能の状態を考慮して修正され、これにより、発話テキストが生成される。

【0084】具体的には、例えば、行動指令情報に含まれるテキストが、「何ですか」である場合において、ロボットの感情の状態が「怒っている」ことを表しているときには、その怒りを表現する「何だよ！」が、発話テキストとして生成される。あるいは、また、例えば、行動指令情報に含まれるテキストが、「やめて下さい」である場合において、ロボットの感情の状態が「怒っている」ことを表しているときには、その怒りを表現する「やめろ！」が、発話テキストとして生成される。

【0085】そして、ステップS5に進み、テキスト生成部31は、発話テキストを対象に、形態素解析や構文解析等のテキスト解析を行い、その発話テキストについて規則音声合成を行うのに必要な情報としての、ピッチ周波数や、パワー、継続時間長等の韻律情報を生成する。さらに、テキスト生成部31は、発話テキストを構成する各単語の発音等の音韻情報も生成する。ここで、ステップS5では、発話テキストの韻律情報として、標準的な韻律情報が生成される。

【0086】その後、テキスト生成部31は、ステップS6において、ステップS5で設定した発話テキストの韻律情報を、ロボットの感情や本能の状態に基づいて修正し、これにより、発話テキストが合成音で出力されるとき感情表現が高められる。

【0087】テキスト生成部31で得られた発話テキストの音韻情報および韻律情報は、規則合成部32に供給され、規則合成部32では、ステップS7において、その音韻情報および韻律情報にしたがい、規則音声合成が行われることにより、発話テキストの合成音のデジタルデータ（合成音データ）が生成される。ここで、規則合成部32でも、規則音声合成の際、感情モデル値や本能モデル値に基づいて、ロボットの感情や本能の状態を適切に表現するように、合成音のポーズの位置や、アクセントの位置、イントネーション等の韻律が変更される。

【0088】規則合成部32で得られた合成音データは、ステップS8において、データバッファ33に供給され、データバッファ33は、規則合成部32からの合成音データを記憶する。

【0089】そして、ステップS9に進み、出力制御部34は、モデル記憶部51に記憶された感情モデル値や

本能モデル値をチェックし、ステップS10に進む。ステップS10では、出力制御部34は、直前のステップS9においてチェックした感情モデル値や本能モデル値に基づき、合成音の出力を中断（停止）するかどうかを判定する。

【0090】ステップS10において、合成音の出力を中断しないと判定された場合、ステップS11に進み、出力制御部34は、データバッファ33から所定量（例えば、1秒分）の合成音データを読み出し、DA変換部35に供給する。DA変換部35では、ステップS12において、出力制御部34からの合成音データがD/A変換され、スピーカ18に供給されて出力される。

【0091】その後、ステップS13に進み、出力制御部34は、データバッファ33に合成音データが記憶されていないかどうか、即ち、データバッファ33が空かどうかを判定する。ステップS13において、データバッファ33が空でなく、まだ、合成音データが記憶されていると判定された場合、ステップS9に戻り、以下、同様の処理が繰り返される。

【0092】従って、この場合は、規則合成部32で生成された合成音が出力され続ける。

【0093】また、ステップS13において、データバッファ33が空であると判定された場合、即ち、データバッファ33に記憶された合成音データの出力が完了した場合、処理を終了する。

【0094】一方、ステップS10において、合成音の出力を中断すると判定された場合、ステップS14に進み、出力制御部34は、データバッファ33をクリアし、ステップS15に進む。ステップS15では、出力制御部34は、必要に応じて、行動決定機構部52に対して、音声合成の対象とするテキストを含む行動指令情報の再出力を要求し、処理を終了する。

【0095】従って、この場合は、合成音の出力が途中で停止される。そして、行動指令情報の再出力の要求が行われた場合には、行動決定機構部52からの行動指令情報の再出力を待って、ステップS1からの処理が行われることにより、合成音の出力が、最初から再開される。

【0096】以上のような音声合成処理において、ステップS10における、合成音の出力を中断するかどうかの判定は、感情モデル値のうちの、例えば、「怒り」を表すものに基づいて行うことができる。即ち、「怒り」の度合いが高い場合に、合成音の出力を停止することができる。

【0097】この場合、合成音が出力されている最中に、ユーザが、ペットロボットを叩く等して、「怒り」の度合いが高くなると、ペットロボットは、即座に合成音の出力を停止する。従って、この場合、ペットロボットが突然黙った状態となることによって、ユーザに対して、ペットロボットが怒った状態にあることを印象づけ

ることができる。

【0098】さらに、「怒り」の度合いが高いが、それほどでもない場合には、出力制御部34において、データバッファ33をクリアした後、行動決定機構部52に対して、行動指令情報の再出力を要求するようにすることができる。

【0099】この場合、「怒り」の度合いが高くなっていることから、上述したように、テキスト生成部31では、「怒り」を表現する発話テキストが生成され、さらに、規則合成部32では、「怒り」を表現する韻律が付された合成音データが生成される。従って、この場合、スピーカ18からは、「怒り」を表す合成音が出力され、その結果、ユーザには、ペットロボットが、いわば怒った口調で言い直しを行ったかのような印象を与えることができる。

【0100】以上のように、ペットロボットの感情の状態に基づいて、合成音の出力を制御するようにしたので、インタラクティブ性の高いペットロボットを実現することができる。

【0101】なお、例えば、ユーザがシステムからの音声出力中に発話を行った場合に、いわゆるバグインに対処する必要から、システムの音声出力を中断する音声対話システムが研究されているが、これは、ユーザによる音声入力を遮らないようにするためであり、感情等のシステムの内部状態に基づいて、システムの音声出力を中断するものではない。従って、このような音声対話システムによれば、ユーザによる音声入力を妨げることを防止することはできるが、本実施の形態におけるペットロボットのように、ユーザとの間のインタラクティブ性を向上させることはできない。

【0102】以上、本発明を、エンターテイメント用のロボット（疑似ペットとしてのロボット）に適用した場合について説明したが、本発明は、これに限らず、例えば、音声合成装置を搭載した対話システムその他に広く適用することが可能である。また、本発明は、現実世界のロボットだけでなく、例えば、液晶ディスプレイ等の表示装置に表示される仮想的なロボットにも適用可能である。

【0103】なお、本実施の形態においては、上述した一連の処理を、CPU10Aにプログラムを実行させることにより行うようにしたが、一連の処理は、それ専用のハードウェアによって行うことも可能である。

【0104】ここで、プログラムは、あらかじめメモリ10B（図2）に記憶させておく他、フロッピーディスク、CD-ROM(Compact Disc Read Only Memory)、MO(Magneto-optical)ディスク、DVD(Digital Versatile Disc)、磁気ディスク、半導体メモリなどのリムーバブル記録媒体に、一時的あるいは永続的に格納（記録）しておくことができる。そして、このようなりムーバブル記録媒体を、いわゆるパッケージソフトウェアとして提供し、ロ

ボット（メモリ10B）にインストールするようにすることができる。

【0105】また、プログラムは、ダウンロードサイトから、デジタル衛星放送用の人工衛星を介して、無線で転送したり、LAN(Local Area Network)、インターネットといったネットワークを介して、有線で転送し、メモリ10Bにインストールすることができる。

【0106】この場合、プログラムがバージョンアップされたとき等に、そのバージョンアップされたプログラムを、メモリ10Bに、容易にインストールすることができる。

【0107】なお、本明細書において、CPU10Aに各種の処理を行わせるためのプログラムを記述する処理ステップは、必ずしもフローチャートとして記載された順序に沿って時系列に処理する必要はなく、並列的あるいは個別に実行される処理（例えば、並列処理あるいはオブジェクトによる処理）も含むものである。

【0108】また、プログラムは、1のCPUにより処理されるものであっても良いし、複数のCPUによって分散処理されるものであっても良い。

【0109】次に、図5の音声合成装置55は、専用のハードウェアにより実現することもできるし、ソフトウェアにより実現することもできる。音声合成装置55をソフトウェアによって実現する場合には、そのソフトウェアを構成するプログラムが、汎用のコンピュータ等にインストールされる。

【0110】そこで、図7は、音声合成装置55を実現するためのプログラムがインストールされるコンピュータの一実施の形態の構成例を示している。

【0111】プログラムは、コンピュータに内蔵されている記録媒体としてのハードディスク105やROM103に予め記録しておくことができる。

【0112】あるいはまた、プログラムは、フロッピー（登録商標）ディスク、CD-ROM、MOディスク、DVD、磁気ディスク、半導体メモリなどのリムーバブル記録媒体111に、一時的あるいは永続的に格納（記録）しておくことができる。このようなリムーバブル記録媒体111は、いわゆるパッケージソフトウェアとして提供することができる。

【0113】なお、プログラムは、上述したようなりムーバブル記録媒体111からコンピュータにインストールする他、ダウンロードサイトから、デジタル衛星放送用の人工衛星を介して、コンピュータに無線で転送したり、LAN、インターネットといったネットワークを介して、コンピュータに有線で転送し、コンピュータでは、そのようにして転送されてくるプログラムを、通信部108で受信し、内蔵するハードディスク105にインストールすることができる。

【0114】コンピュータは、CPU(Central Processing Unit)102を内蔵している。CPU102には、バス1

01を介して、入出力インタフェース110が接続されており、CPU102は、入出力インタフェース110を介して、ユーザによって、キーボードや、マウス、マイク等で構成される入力部107が操作等されることにより指令が入力されると、それにしたがって、ROM(Read Only Memory)103に格納されているプログラムを実行する。あるいは、また、CPU102は、ハードディスク105に格納されているプログラム、衛星若しくはネットワークから転送され、通信部108で受信されてハードディスク105にインストールされたプログラム、またはドライブ109に装着されたリムーバブル記録媒体111から読み出されてハードディスク105にインストールされたプログラムを、RAM(Random Access Memory)104にロードして実行する。これにより、CPU102は、上述したフローチャートにしたがった処理、あるいは上述したブロック図の構成により行われる処理を行う。そして、CPU102は、その処理結果を、必要に応じて、例えば、入出力インタフェース110を介して、LCD(Liquid Crystal Display)やスピーカ等で構成される出力部106から出力、あるいは、通信部108から送信、さらには、ハードディスク105に記録等させる。

【0115】なお、本実施の形態では、行動決定機構部52が生成するテキストから合成音を生成するようにしたが、本発明は、あらかじめ用意されたテキストから合成音を生成する場合にも適用可能である。さらに、本発明は、あらかじめ録音してある音声データを編集して、目的とする合成音を生成する場合にも適用可能である。

【0116】また、本実施の形態では、ペットロボットの感情の状態に基づいて、合成音の出力を制御するようにしたが、合成音の出力は、その他、例えば、本能や成長その他のペットロボットの内部状態に基づいて制御することが可能である。

【0117】

【発明の効果】以上の如く、本発明の音声合成装置および音声合成方法、並びにプログラムによれば、情報処理装置の制御にしたがって、合成音が生成される一方、情報処理装置の内部状態に基づいて、合成音の出力が制御

される。従って、インタラクティブ性の高い合成音の出力を行うことが可能となる。

【図面の簡単な説明】

【図1】本発明を適用したロボットの一実施の形態の外観構成例を示す斜視図である。

【図2】ロボットの内部構成例を示すブロック図である。

【図3】コントローラ10の機能的構成例を示すブロック図である。

【図4】音声認識部50Aの構成例を示すブロック図である。

【図5】音声合成部55の構成例を示すブロック図である。

【図6】音声合成部55による音声合成処理を説明するフローチャートである。

【図7】本発明を適用したコンピュータの一実施の形態の構成例を示すブロック図である。

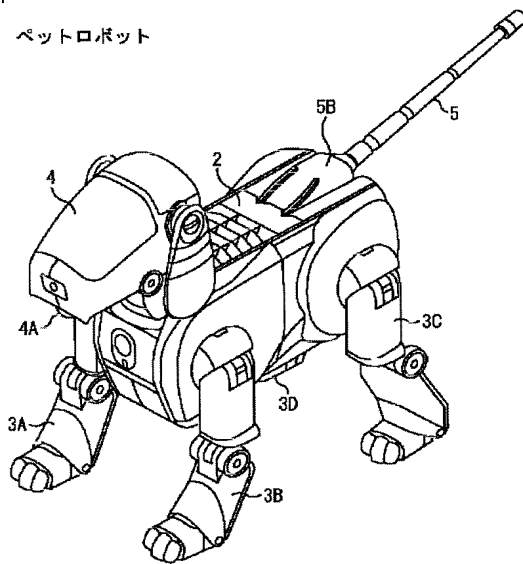
【符号の説明】

1 頭部ユニット, 4A 下顎部, 10 コントローラ, 10A CPU, 10B メモリ, 15 マイク, 16 CCDカメラ, 17 タッチセンサ, 18 スピーカ, 21 AD変換部, 22 特徴抽出部, 23 マッチング部, 24 音響モデル記憶部, 25 辞書記憶部, 26 文法記憶部, 27 音声区間検出部, 31 テキスト生成部, 32 規則合成部, 33 データバッファ, 34 出力制御部, 35 DA変換部, 36 辞書記憶部, 37 生成用文法記憶部, 38 音素片記憶部, 50 センサ入力処理部, 50A 音声認識部, 50B 画像認識部, 50C 圧力処理部, 51 モデル記憶部, 52 行動決定機構部, 53 姿勢遷移機構部, 54 制御機構部, 55 音声合成部, 101 バス, 102 CPU, 103 ROM, 104 RAM, 105 ハードディスク, 106 出力部, 107 入力部, 108 通信部, 109 ドライブ, 110 入出力インタフェース, 111 リムーバブル記録媒体

【図1】

図1

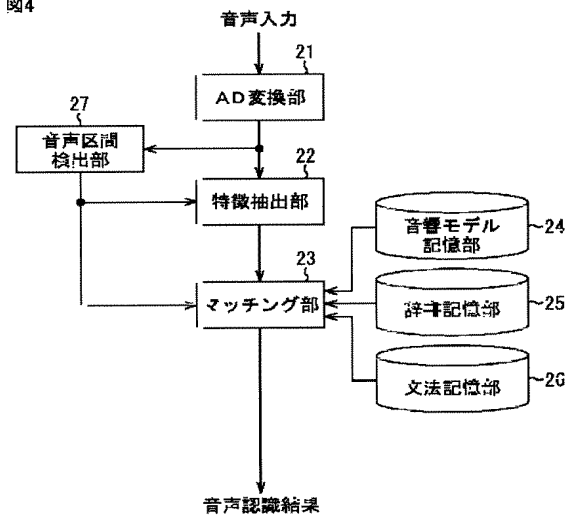
ペットロボット



ペットロボットの外観構成

【図4】

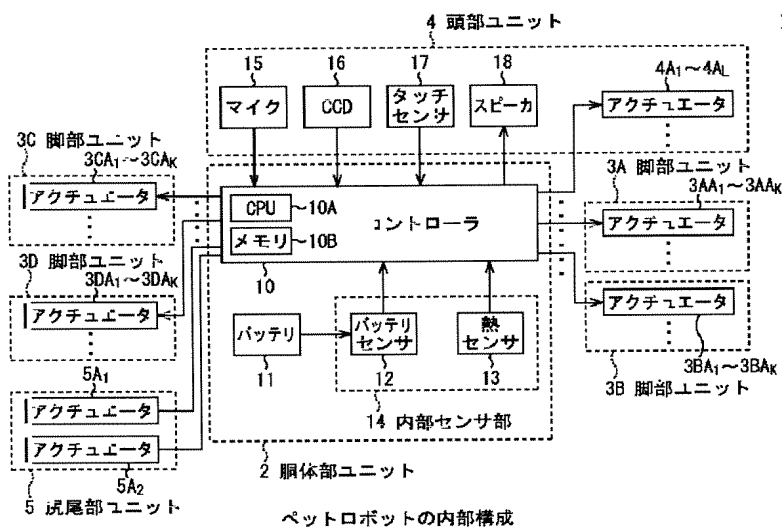
図4



音声認識部 50A

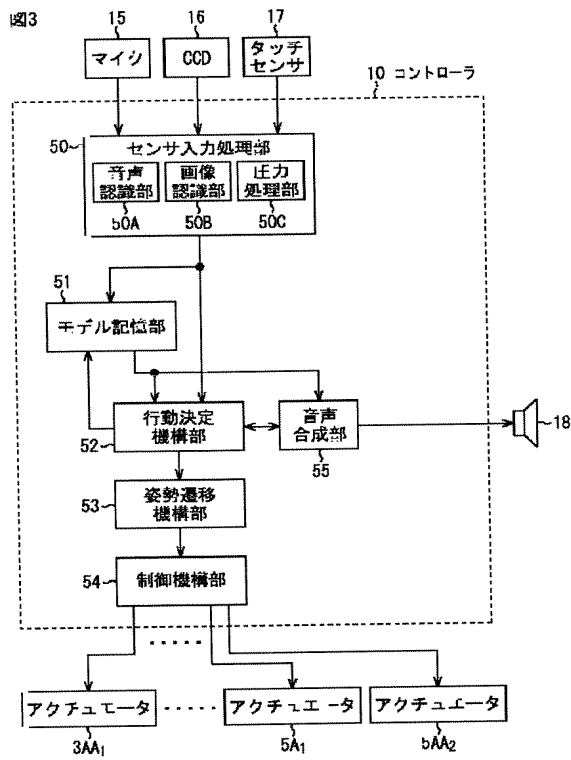
【図2】

図2

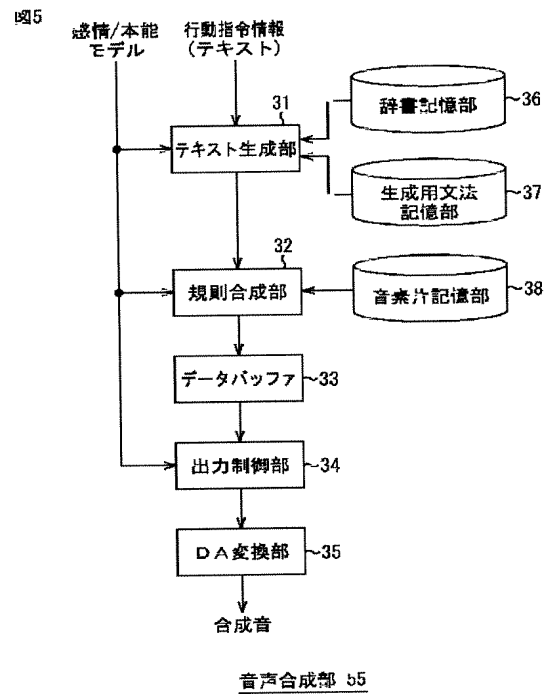


ペットロボットの内部構成

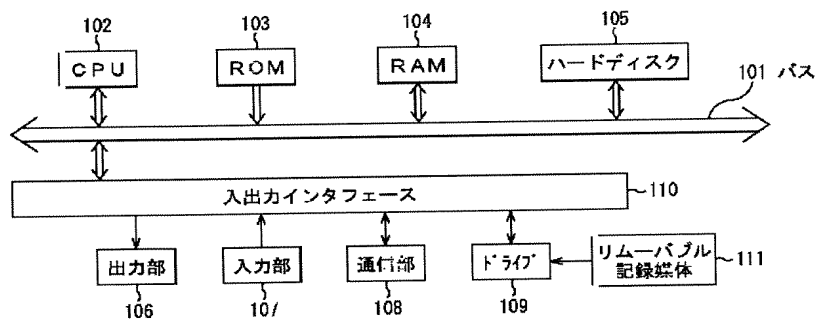
【図3】



【図5】

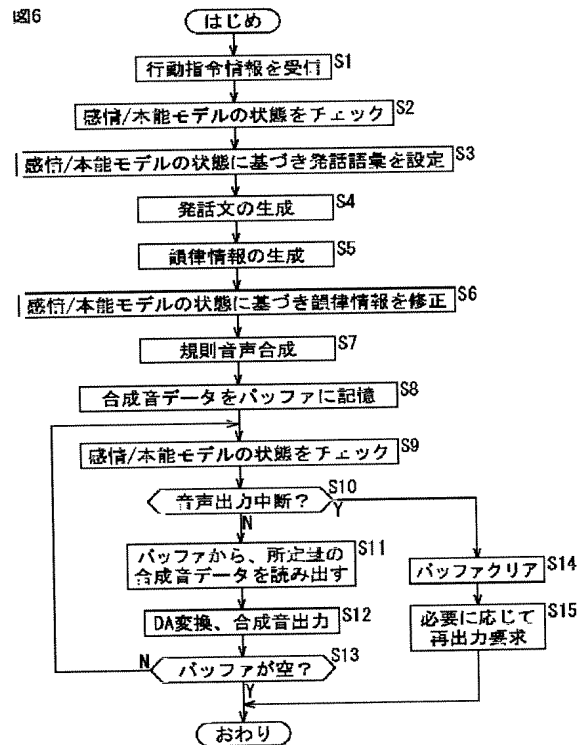


【図7】



コンピュータ

【図6】



フロントページの続き

(72)発明者 山崎 信英
東京都品川区北品川6丁目7番35号 ソニ
ー株式会社内

(72)発明者 狩谷 真一
東京都品川区北品川6丁目7番35号 ソニ
ー株式会社内

(72)発明者 藤田 八重子
東京都品川区北品川6丁目7番35号 ソニ
ー株式会社内

Fターム(参考) 2C150 CA01 CA02 CA04 DA05 DA24
DA25 DA26 DA27 DA28 DF03
DF04 DF33 ED42 ED52 EF03
EF07 EF09 EF13 EF16 EF23
EF29 EF34 EF36
3C007 AS36 CS08 KS10 MT14 WA04
WA14 WB16 WB28 WC30
5D045 AA08 AA09 AB11